REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-98-

0092

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
		Final (15 June 1994 – 14 June 1997)

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
(FY94) Supplement to Exploiting Chaos in Oversampled A/D Converters	F49620-94-1-0359
6. AUTHORS Professor Avideh Zakhor	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California Berkeley, CA 94720	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/NM 110 Duncan Avenue, Room B-115 Bolling Air Force Base, DC 20332-8080	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The grant supported work in three-dimensional image representation and reconstruction. Multi-level information is combined to enable effective representation of shading and occlusion. The modeling process fits in well with CCD camera array sensing, and multi-channel sampling permits improved motion estimates between frames in a video sequence. The usual redundancy in 3-D scene representation can be shown to allow meaningful exploitation by encoding the redundant distance, depth and intensity parameters. Significant performance improvements were demonstrated over usual bilinear and cubic B-spline algorithms.

14. SUBJECT TERMS	15. NUMBER OF PAGES
three-dimensional image representation and reconstruction	
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

Final Technical Report

Air Force Office of Scientific Research Grant No. F49620-94-1-0359

Professor Avideh Zakhor, Principal Investigator

August 1, 1994 - June 14, 1997

Final Report

Avideh Zakhor
Department of Electrical Engineering and Computer Sciences
University of California
Berkeley, CA 94720

1 Introduction

We have studied, developed and examined solutions to the problems of 3-D scene and image sequence representation and new view synthesis. We also have developed a video resolution enhancement algorithm. This report summarizes our efforts in each area.

2 Depth-Based Representations for Image Reconstruction and New View Synthesis

The problems of image sequence compression and new view synthesis have both received a lot of attention recently. In the former case, it is desired to compactly represent the original image set by exploiting redundancy and correlation. This issue is particularly important in applications of storage and transmission. In contrast, the goal of new view synthesis is to generate arbitrary viewpoints of a given scene primarily for visualization purposes. Notice that there exists a tradeoff between representation size and the quality of the synthesized images: As more views of the scene are added to the representation, the image quality increases as does the representation size. Hence, an interesting problem is to consider both problems at once; that is, construct a compact representation which reconstructs the original images and synthesizes new views.

We developed two depth-based representations to address these problems. The first approach involves several so-called *reference frames* for which depth and intensity information are both defined. New views are generated by warping the reference intensity and depth data in a manner similar to view interpolation techniques [3, 6, 14, 1, 2, 11, 5]. The second approach integrates all available information with respect to a single reference frame akin to layered representations [7, 21, 19, 22]. The representation then consists of a multivalued array of depth and intensity values which overcomes occlusions and redundancy [4]. These depth-based representations both assume the given image sequences arise from a static 3-D scene captured by a moving camera restricted to the x-y plane. Note that the exact motion of the camera is unknown *a priori* and will be estimated.

2.1 Depth Estimation and Synthesis

Given an image sequence, it seems intuitive to compute depth pairwise between the reference frame and each of its neighbors to generate local "depth maps". Since every frame is related by a planar translation, depth estimation can be accomplished by 1-D correspondence matching along the parallel epipolar lines. In [1], the l_2 norm of intensity error is minimized over possible depth values using adaptive neighborhoods $\mathcal{N}$:

$$\min_d \left\{ \sum_{(u,v)} \sum_{\in \mathcal{N}} \| I_k(u,v) - I_i(u',v') \|^2 \right\} \tag{1}$$

where predicted coordinates (u', v') and disparity d are related to a candidate motion vector (m, n) by

$$u' = u + m \tag{2}$$

$$v' = v + n \tag{3}$$

$$d = \sqrt{m^2 + n^2} = \frac{f}{z}\sqrt{b_x^2 + b_y^2}. \tag{4}$$

While pairwise matching leads to reasonable depth results, multiframe approaches perform even better by reducing ambiguity and increasing accuracy when camera motion is known. To compute depth for a particular frame, a variant of Okutomi and Kanade's multiple-baseline algorithm is used [16]. The approach consists of finding the inverse depths that minimize the sum of component intensity errors. More precisely, suppose there are M images denoted by $I_i(\cdot,\cdot)$ and let $k \in 1, 2, \ldots M$ be the reference frame. Then, the goal is to compute inverse depth ζ for every desired point with the following expression

$$\min_\zeta \left\{ \sum_{i \neq k}^{M} \sigma_i \left(\sum_{(u,v)} \sum_{\in \mathcal{N}} \| I_k(u,v) - I_i(u',v') \|^2 \right) \right\} \tag{5}$$

where $\mathcal{N}$ is a local neighborhood around the pixel of interest, σ_i indicates the influence of frame i, and (u', v') are the predicted image coordinates. For planar translation, they are given by

$$u' = u - f b_{xi} \zeta \tag{6}$$

$$v' = v - f b_{yi} \zeta \tag{7}$$

Assuming the baselines (b_{xi}, b_{yi}) are known *a priori* or else computed, one can proceed to estimate the inverse depths ζ using Eqn (5) for all desired points in the frame.

Our implementation of the multiple-baseline algorithm differs from Okutomi and Kanade's in several ways. First, adaptive neighborhood sizes for $\mathcal{N}$ are employed to improve estimation in low-textured regions. The neighborhood is automatically adjusted according to the

local variance of neighboring intensities [5]. Next, instead of normalizing the largest baseline to be 1, one of the shorter baselines is considered to have unity baseline. This feature permits wider baselines to be included without drastically increasing computational time.

Because wider baselines may be used, occlusions in the scene will pose a larger problem in multiframe matching. The effects of occlusions are mitigated by the addition of σ_i in Eqn (5) which will be on only for the frames in which the point is visible [4].

2.2 View Synthesis

Once a dense depth map has been computed using either pairwise or multiframe matching, it is relatively straightforward to warp the reference information to synthesize new views of the scene. The procedure consists of regarding the depth map as a deformable mesh of quadrilateral patches [5]. Vertices of each patch are warped by the appropriate transformation. For reconstruction of the original images, the transformation is simply

$$u' = u + fb_x/Z \tag{8}$$

$$v' = v + fb_y/Z \tag{9}$$

where f is the focal length, (b_x, b_y) is the amount of planar translation, and Z is the depth corresponding to point (u, v). Alternatively, off-plane views may be obtained by using the transformation

$$u' = f\frac{r_{1,1}X + r_{1,2}Y + r_{1,3}Z + \Delta x}{r_{3,1}X + r_{3,2}Y + r_{3,3}Z + \Delta z} \tag{10}$$

$$v' = f\frac{r_{2,1}X + r_{2,2}Y + r_{2,3}Z + \Delta y}{r_{3,1}X + r_{3,2}Y + r_{3,3}Z + \Delta z}. \tag{11}$$

The interior is rendered using a traditional 2-D scan-line algorithm and Z-buffering to ensure the proper depth ordering [8]. Patches which transcend depth edges are not rendered since they may lead to "smearing" [5]. In the end, it is possible for the final image to contain "holes" which correspond to slight inaccuracies in the estimated depth or to regions unseen in the original frames.

To illustrate this synthesis procedure, consider frame 35 from the Mug sequence in Section 2.3 as shown in Figure 1 (a). Pairwise matching is performed between frame 35 and every one of its neighbors. The local depth maps are then combined to form Figure 1 (b). Figure 1 (c) is the result of warping every pixel according to its depth to synthesize a translated virtual camera.

2.3 View Interpolation

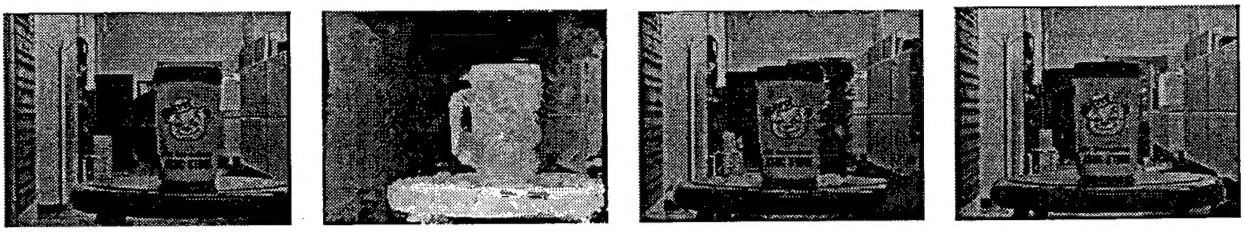
It is clear from Section 2.2 that novel views of the scene may be synthesized quite accurately and easily from a single reference intensity-depth pair. Further improvements can be made

(a) (b) (c)

Figure 1: *Example of synthesizing new view from a single reference pair: (a) intensity image frame 35 of Mug; (b) corresponding depth map; and (c) synthesized view. The depth map is quantized to 256 gray levels where the depth is inversely related to the brightness. Note that depth has also been histogram equalized to show the contrast between the object and the surrounding background. Holes shown in red correspond to regions that become uncovered.*

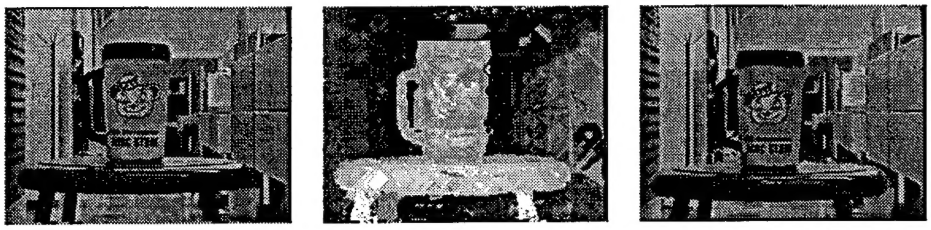
by introducing a second or multiple reference pairs. Hence, our first proposed representation consists of employing multiple reference pairs. One may derive this representation using the techniques described in Section 2.1 in the following steps:

1. *Compute dense depth for every reference frame.*
2. *Estimate motion between reference frames.*
3. *Discard neighboring frames to form representation.*
4. *Generate view estimates and combine to form desired view.*



(a) (b) (c) (d)

Figure 2: *Reconstruction of horizontal view from reference frame 35 and 65 of Mug: (a) intensity image frame 65 of Mug; (b) corresponding depth map; (c) view estimate using only reference frame 65; and (d) reconstructed view combining view estimates.*



(a) (b) (c)

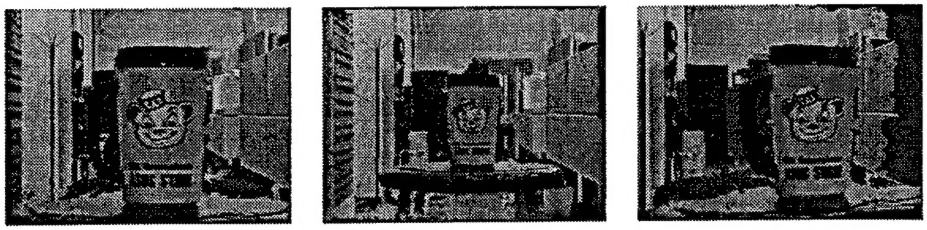
Figure 3: *Reconstruction of vertical view from reference frame 35 of Mug and frame 37 of Mug2: (a) intensity image frame 37 of Mug2; (b) corresponding depth map; and (c) reconstructed view.*

The above steps are applied to a real-world scene filmed by a cam-corder undergoing unknown horizontal translation at two different elevations. The two sequences, known as Mug and Mug2, were digitized to 320×240 and subsampled temporally to obtain eighteen Mug frames and seven Mug2 frames. Three frames, frames 35 and 65 from Mug and frame 37 from Mug2, were chosen to serve as reference frames; Figures 1, 2, and 3 show these reference pairs, respectively.

Using reference frames 35 and 65, the midpoint view along the same horizontal trajectory is chosen to be reconstructed. Using only reference frame 35 or 65 leads to the view estimates shown in Figure 1 (c) and Figure 2 (c), respectively. Since the holes in the view estimates do not overlap, one would expect improved results after combining the view estimates. As shown in Figure 2 (d), the combined result quality is good for the most part. The horizontal edges, e.g. top of the door, top of the mug, specularities in front of the stool, and the drawers, have been reconstructed quite well. The proposed approach takes care of problems in occluded regions; there are only a few errors to the right of the mug and near the mug handle. These artifacts arise because the depth edges were not localized perfectly.

To generate a view not originally scanned by the cam-corder, reference frames 35 and 37 are used to synthesize the midpoint on the vertical trajectory relating the two views; the result is given in Figure 3 (c). The image is a reasonable estimate of the desired view. As before, the most troublesome region in the image lies inside the handle of the mug.

More interesting views not necessarily confined to the x-y plane may be reconstructed with this representation. For instance, the viewpoint of a camera translated toward the scene can also be rendered quite easily; the resulting image is given in Figure 4 (a). Note that this view differs from a simple "zoom-in" since the latter requires only a larger focal length and it does not uncover occluded regions. The two regions above the stool are marked red because none of the reference frames has information about what lies behind the stool in the scene. Figure 4 (b) shows the view translated away from the scene with the uncovered



(a) (b) (c)

Figure 4: *Examples of synthesized views using multiple reference frames: (a) translation toward scene; (b) translation away from scene; and (c) arbitrary rotation and translation.*

regions marked accordingly. Finally, Figure 4 (c) shows an oblique view of the scene taken by rotating the camera 10° clockwise and translating along both the x and z axes. The quality of the synthesized image is quite good given the amount of uncovered regions.

2.4 Multivalued Representation

In representing a 3-D scene, it is common for the images to be very similar and to exhibit a lot of redundancy. This fact is especially true when the images come from arbitrary translational motion in the x-y plane since the depth of scene points remains fixed in all the images. One possible compact representation for this case would involve remapping all visible information with respect to one particular frame. We thus consider exploiting the redundancy to form a multivalued representation (MVR) of depth and intensity. The MVR separates information into levels of occlusion and can easily handle points occluded from reference viewpoint.

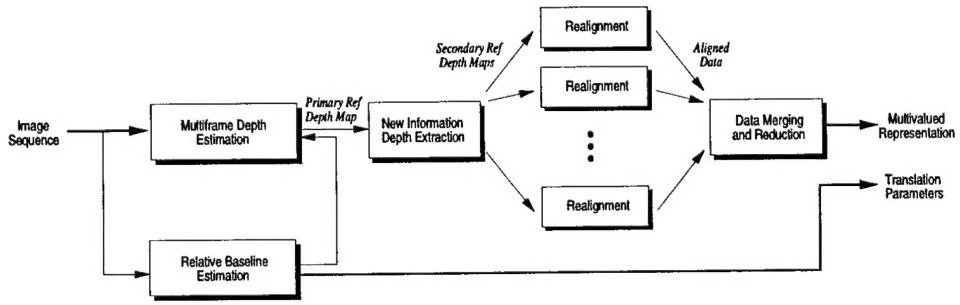


Figure 5: *Block diagram for the multivalued representation.*

To build a MVR from a set of images, one first selects a single frame, denoted as the *primary reference frame* or PRF, for which the representation is defined. As diagrammed in Figure 5, the following steps are then performed:

1. *Estimate motion parameters between PRF and each neighbor.*
2. *Calculate dense depth for PRF using multiframe algorithm.*
3. *Compute depth for new information in other frames.*
4. *Fit piecewise 3-D surfaces through depth maps.*
5. *Merge and reduce data to produce final MVR.*

The final result consists of a multivalued array of intensities and depths corresponding to the primary reference frame. Notice that the information contained in the MVR consists of the union of intensity and depth that can be extracted from the original image data.

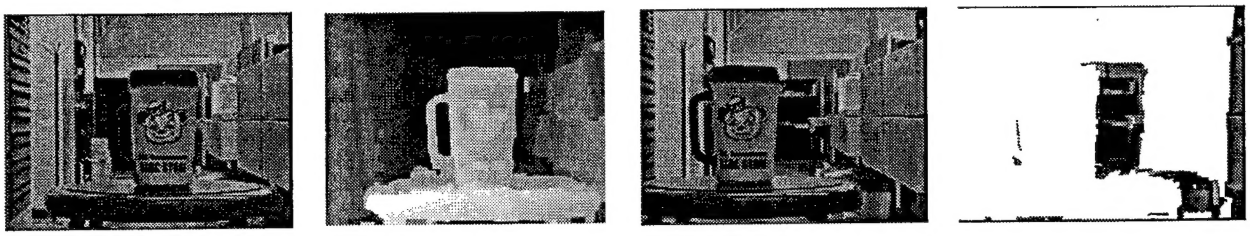
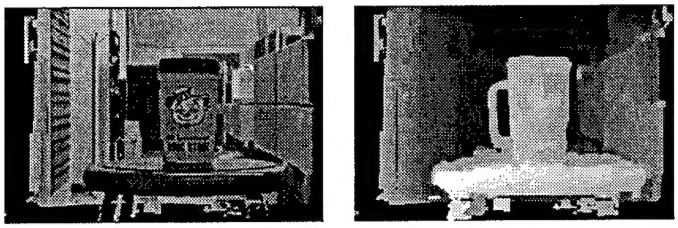


Figure 6: *Example of estimating new information: (a) intensity PRF 50 of Mug; (b) depth PRF 50; (c) intensity frame 21 of Mug; and (d) new information in frame 21 wrt frame 50. As expected, the algorithm identifies the cubicle located behind the mug as well as the right border of the image, both obscured from view in frame 50.*

As before, we consider the Mug and Mug2 sequences, where only nine frames of Mug and four from Mug2 are used. Frame 50, shown in Figure 6 (a), is selected as the primary reference frame for the representation. Using the multiframe algorithm leads to the depth map found in Figure 6 (b). Notice the accuracy of the estimated depths especially the descending walls. The synthesis techniques of Section 2.2 may be applied to this depth map to obtain an estimate of, say, frame 21. If this view estimate is compared with the original image (see Figure 6 (c)), one can easily extract the new information contained in frame 21 with respect to the PRF as shown in Figure 6 (d).

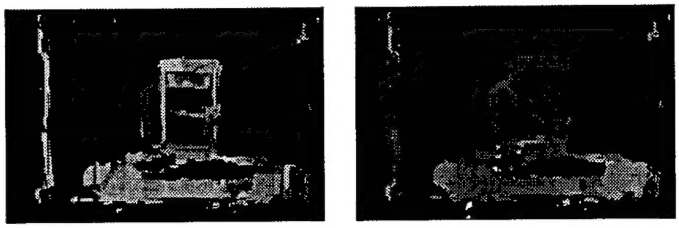
Applying the above algorithm, dense depth corresponding to the points visible from the PRF as well as points occluded in this frame are recovered. The intensity and depth information in level 0 are shown in Figures 7 (a) and (b). Points shown in blue correspond to regions without intensity and depth. The shape of the mug and the stool have been recovered quite well. Notice that the left and right sides descend in depth as expected. Also, the dimensions of the original image have been expanded and the points seen along the borders have been recovered. Even the legs of the stool have been extended.



(a) (b)

Figure 7: *Recovered information for level 0 of the MVR: (a) intensity and (b) depth. The depth is quantized to 256 gray levels where the depth is inversely related to the brightness. Note that depth has also been histogram equalized to show the contrast between the object and the surrounding background.*

Figures 8 (a) and (b) show the recovered information in the second level of the MVR. Most of the information corresponds to points that are located behind the mug. The cubicle and the wall are both recovered from behind the mug since they were seen in some of the original images. Moreover, the ground obscured by the stool is revealed in this level. By filling in points from level 0 as in Figure 9, it appears that the mug and most of the stool have been removed. Notice that the bottom portion of the legs and part of the stool remain since the regions behind them were occluded in the original images.



(a) (b)

Figure 8: *Recovered information for level 1 of the MVR: (a) intensity and (b) depth. The cubicle located behind the mug was recovered in both intensity and depth domains. Also the wall behind the mug handle and the floor behind the stool are revealed.*

The reconstruction techniques of Section 2.2 are applied to generate the original images. As an example, Frame 21 has been reconstructed in Figure 10 (a). Notice that the reconstructed quality is quite good. Similar quality is obtained in the other reconstructed images as seen in Figures 10 (b)–(d). The average PSNR for reconstructed images is 30.707 dB.

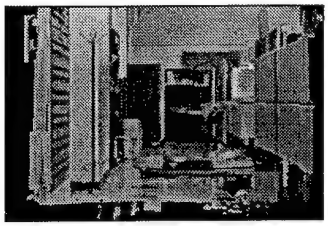
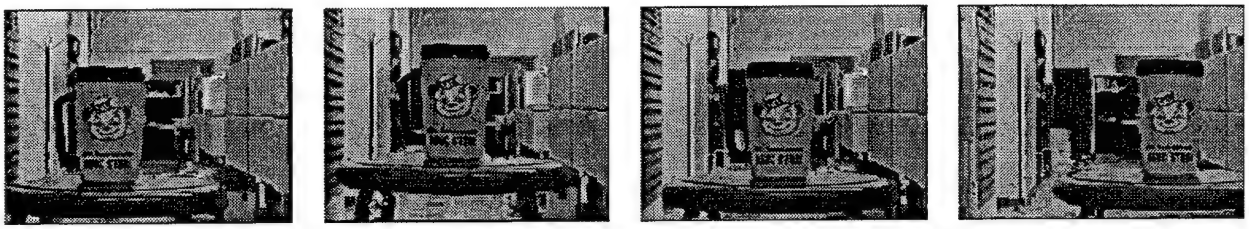


Figure 9: *Points from level 1 are combined with points from level 0 to put the representation in context.*



(a) (b) (c) (d)

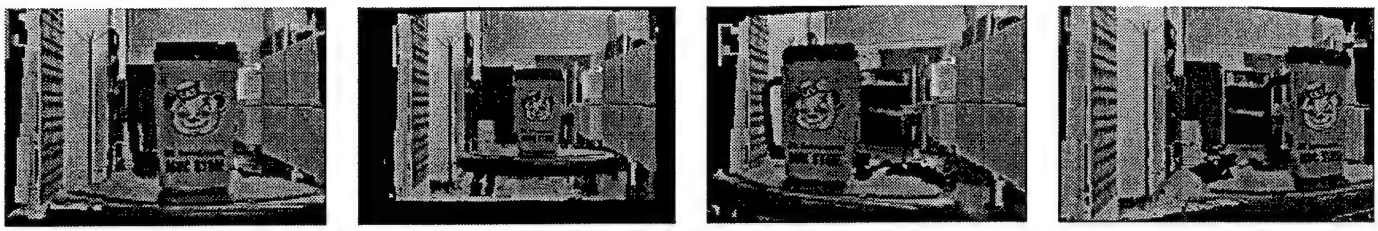
Figure 10: *Examples of reconstructed views using MVR: (a) frame 21; (a) frame 37; (a) frame 40; and (d) frame 80.*

Synthesized views of the scene may be generated in a similar manner. Translations toward and away from the scene are given in Figures 11 (a) and (b), respectively. Figures 11 (c) and (d) show the virtual camera undergoing arbitrary motion. Despite an increase in the number of artifacts for these views, the resulting images are reasonable and provide a convincing sense of depth.

3 Multiframe Spatial resolution Enhancement

Much recent research has focused on using signal processing to enhance the spatial resolution of images. The problem is one of image interpolation, where unknown pixels must be determined by using the constraints provided by the known pixel values. Since single frame interpolation methods are inherently limited by the amount of data available to constrain the solution, multiframe methods have been proposed to add constraints to the problem. Multiframe approaches depend on a motion estimation stage which enables combination of several low resolution frames.

We developed a two-stage technique to produce each high resolution frame of the video sequence. The general technique is shown in Figure 12. First, we use a registration technique



(a) (b) (c) (d)

Figure 11: *Examples of synthesized views using MVR: (a) translation toward scene; (b) translation away from scene; (c) and (d) arbitrary rotation and translation.*

to determine a dense set of subpixel accuracy candidate motion vectors for several low resolution frames relative to a reference low resolution frame. Choosing a set of motion vectors per pixel instead of a single estimate is motivated by our model of the imaging process and the knowledge that perfect motion estimation is not possible in general. Next, we incorporate a sequence of low resolution frames into an initial high resolution frame estimate using the motion estimation results. Then, we apply an iterative algorithm which uses the low resolution pixel intensity and motion estimation constraints to improve the quality of our initial high resolution estimate.

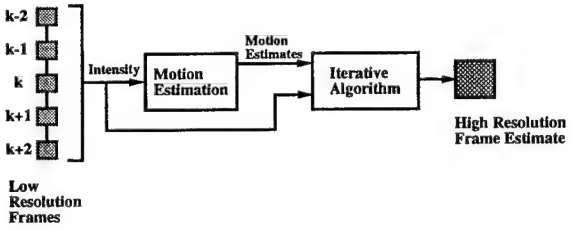


Figure 12: System diagram of enhancement algorithm.

We model the imaging process based on the operation of a CCD sensor array of a camera. Since a low resolution camera simply has a larger detector area than a high resolution camera, we model the imaging process as a spatial averaging and subsampling. If the high resolution image contains any frequencies above the Nyquist frequency then there will be aliasing in the low resolution image. This aliasing is, of course, what we wish to remove with our multiframe enhancement technique. The generalized multichannel sampling theorem [17] tells us that we can do this, provided that we have perfect motion estimates between frames. The aliasing and blurring caused by the imaging process, however, makes such perfect motion estimation impossible.

3.1 Motion Estimation

As in previous methods, we perform pairwise subpixel accuracy motion estimation for each pixel of each low resolution frame relative to a reference frame. Since our experiments will deal with enhancement by a factor of two in each dimension, we perform half-pixel accuracy motion estimation. The accuracy of motion estimation may be increased for larger enhancement factors. A novelty of our technique involves how we overcome accuracy limitations caused by the imaging process. Unlike in previous attempts, our approach is to save a small set of candidate motion vectors for each pixel in each frame, instead of limiting ourselves to a single motion estimate per pixel. This set of candidate motion estimates is obtained by saving all motion estimates within a small threshold of the "best" estimate in terms of minimum MSE. We do not simply choose the minimum MSE motion estimates because the blurring and aliasing of the imaging process sometimes cause the correct motion estimate to have a non-minimum MSE.

A second motion estimation feature we utilize to increase accuracy is using the chrominance components, in addition to the luminance component, to compute our motion vectors. Our experiments verify previous studies [9] that using color components yields significant improvement over luminance only motion estimation.

3.2 Initial Estimate

After the motion estimation stage, we use the resulting sets of candidate motion vectors to combine the low resolution intensity frames, obtaining an initial high resolution frame estimate by mapping low resolution pixels to high resolution ones. Since we have a set of motion estimates for each pixel in each low resolution frame, instead of just one estimate, several scenarios can arise. In the case where exactly one low resolution pixel maps to a high resolution pixel, we keep this pixel and the corresponding motion vector. Another possibility is multiple motion vectors, either from a single frame or from several different frames, mapping several low resolution pixels to a single high resolution pixel. In this case, we choose the pixel and motion vector with the smallest MSE. Selecting low resolution pixel intensities and corresponding motion vectors in this manner enables us to reduce all the sets of candidate motion vectors from the different frames into a final set of only one intensity and one motion-vector for each high resolution pixel. A final possibility is the existence of "holes" where some high resolution pixels have no low resolution pixels mapped to them by any of the multiple motion vectors. This possibility is the worst case because we have no intensity or motion information about the corresponding high resolution pixel. Fortunately, using multiple motion estimates per pixel greatly reduces the number of holes compared to using only a single vector per pixel. Any existing holes in the initial high resolution frame estimate may be filled by a simple interpolation technique. The iterative algorithm will modify all of these initial intensity values as described below.

3.3 Iterative Enhancement Algorithm

Our iterative enhancement technique, based on Landweber's algorithm [13], uses the combined motion estimates along with the low resolution intensity frames to iteratively modify the initial estimate. The process adjusts the high resolution estimate, according to the constraints provided by the low resolution intensities and the motion estimates, to converge upon the final high resolution frame estimate. The iterative technique relies on the fact that our only information about the original high resolution frame is embedded in the sequence of low resolution frames. At each iteration we apply a simulated imaging process to the current high resolution estimate to obtain a set of simulated low resolution frames. We then compare the set of original low resolution frames with this set of simulated frames and modify the high resolution estimate in such a way as to make the simulated set of frames more closely match the original ones at the next iteration. This approach quickly converges to the final high resolution estimate.

3.4 Results

To test our approach and compare it with other methods, we applied our technique to the *Foreman* sequence. We also applied bilinear interpolation and cubic B–spline interpolation to compare with our method. Figure 13 shows the qualitative results and Figure 14 shows the quantitative results for the *Foreman* sequence. The average PSNR is 27.78 dB when using bilinear interpolation, 28.71 using cubic B-spline interpolation, and 31.29 dB using our multiframe algorithm, i.e., a gain of 3.51 dB and 2.58 dB over bilinear and cubic B-splines algorithms.

References

[1] N. L. Chang, "View reconstruction from uncalibrated cameras for three-dimensional scenes," Master's thesis, University of California at Berkeley, 1994.

[2] N. L. Chang and A. Zakhor, "Arbitrary view generation for three-dimensional scenes from uncalibrated video cameras," in *Proceedings of ICASSP*, pp. 2455–2458, Detroit, MI, 8–12 May 1995.

[3] N. L. Chang and A. Zakhor, "Intermediate view reconstruction for three-dimensional scenes," in *Proceedings of ICDSP*, vol. 2, pp. 636–641, Nicosia, Cyprus, 14–16 July 1993.

[4] N. L. Chang and A. Zakhor, "Multivalued representations for image reconstruction and new view synthesis." In preparation.

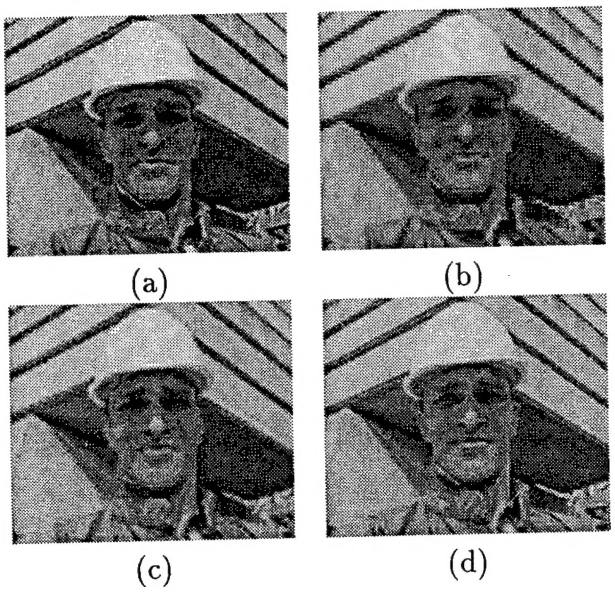


Figure 13: *Foreman* video sequence: (a) Original high resolution frame, (b) low resolution frame, (c) cubic B–spline interpolation, (d) multiframe estimate.

[5] N. L. Chang and A. Zakhor, "View generation for three-dimensional scenes from video sequences," *IEEE Trans. on Image Proc.*, vol. 6, no. 4, Apr. 1997. To appear.

[6] S. E. Chen and L. Williams, "View interpolation for image synthesis," in *Proceedings of SIGGRAPH*, pp. 279–288, New York, NY, 1–6 Aug. 1993.

[7] T. Darrell and A. Pentland, "Robust estimation of a multi-layered motion representation," in *IEEE Workshop on Visual Motion*, pp. 173–178, Princeton, NJ, 7–9 Oct. 1991.

[8] J. D. Foley, A. van Dam, et al., *Introduction to Computer Graphics.* Addison-Wesley, 1994.

[9] H. Harasaki and A. Zakhor, "Motion compensation using color component signals." Presented at the IEEE Workshop on Visual Signal Processing and Communications, June 1991.

[10] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 231–239, May 1991.

[11] T. Kanade, P. J. Narayanan, and P. W. Rander, "Virtualized reality: Concepts and early results," in *IEEE Workshop on Representation of Visual Scenes*, pp. 69–76, Cambridge, MA, June 24 1995.

[12] T. Komatsu, T. Igarashi, K. Aizawa, and T. Saito, "Very high resolution imaging scheme with multiple different-aperture cameras," *Signal Processing: Image Communication*, vol. 5, pp. 511–526, December 1993.

[13] L. Landweber, "An iterative formula for fredholm integral equations of the first kind," *American Journal of Mathematics*, vol. 73, pp. 615–624, 1951.

[14] S. Laveau and O. Faugeras, "3-D scene representation as a collection of images and fundamental matrices," Tech. Rep. 2205, INRIA, Feb. 1994.

[15] H. Maître and W. Luo, "Using models to improve stereo reconstruction," *IEEE Trans. on Patt. Anal. Mach. Intell.*, vol. 14, no. 2, pp. 269–277, Feb. 1992.

[16] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Trans. on Patt. Anal. Mach. Intell.*, vol. 15, no. 4, pp. 353–363, Apr. 1993.

[17] A. Papoulis, "Generalized sampling expansion," *IEEE Transactions on Circuits and Systems*, vol. 24, pp. 652–654, November 1977.

[18] N. R. Shah, "Multiframe algorithm for spatial resolution enhancement of color video sequences," Master's thesis, University of California at Berkeley, December 1995.

[19] H. S. Sawhney and S. Ayer, "Compact representations of videos through dominant and multiple motion estimation," *IEEE Trans. on Patt. Anal. Mach. Intell.*, vol. 18, no. 8, pp. 814–830, Aug. 1996.

[20] R. R. Schultz and R. L. Stevenson, "Improved definition video frame enhancement," in *International Conference on Acoustics, Speech and Signal Processing*, vol. 4, (Detroit, MI), pp. 2169–2172, IEEE, May 1995.

[21] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. on Image Proc.*, vol. 3, no. 5, pp. 625–638, Sept. 1994.

[22] Y. Weiss and E. H. Adelson, "A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models," in *Proceedings of CVPR*, pp. 321–326, San Francisco, CA, 18–20 June 1996.

[23] Z. Zhang, R. Deriche, et al., "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, no. 1–2, pp. 87–119, Oct. 1995.

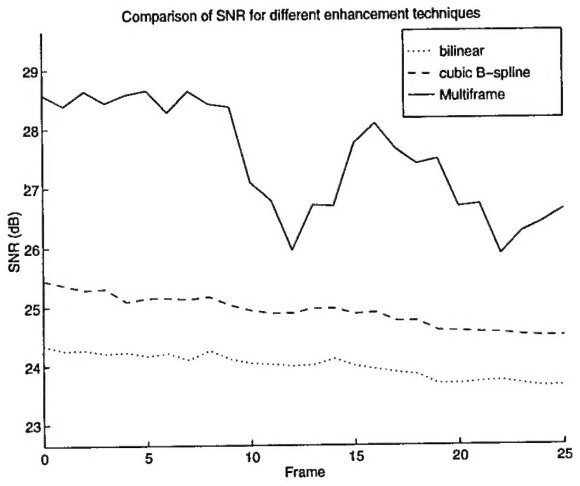


Figure 14: SNR plots of various techniques applied to the *Foreman* sequence.